IDIAP RESEARCH REPORT



EFFICACY AND EFFECTIVENESS OF ROBOT-ASSISTED THERAPY FOR AUTISM SPECTRUM DISORDER: FROM LAB TO REALITY

Daniel David Paul Baxter Tony Belpaeme
Erik Billing Hoang-Long Cao Haibin Cai
Anamaria Ciocan Cristina Costescu
Daniel Hernandez Garcia Pablo Gomez Esteban
James Kennedy Honghai Liu Silviu Matu
Alexandre Mazel Mihaela Selescu
Emmanuel Senft Serge Thill Bram Vanderborght
David Vernon Tom Ziemke

Idiap-RR-06-2026

NOVEMBER 2026

Centre du Parc, Centre du Parc, Rue Marconi 19, CH - 1920 Martigny
T +41 27 721 77 11 F +41 27 721 77 12 info@idiap.ch www.idiap.ch

This is an unproofed preprint. Citations should refer to the published version:

David et al. (2025). Efficacy and effectiveness of robot-assisted therapy for autism spectrum disorder: From lab to reality. *Science Robotics*, 10(109), eadl2266. DOI: 10.1126/scirobotics.adl2266.

Title:

Efficacy and Effectiveness of Robot-Assisted Therapy for Autism Spectrum Disorder: From Lab to Reality

Authors:

Daniel David[1, 2], Paul Baxter[3], Tony Belpaeme[4], Erik Billing[5], Haibin Cai[6], Hoang-Long Cao[7,8], Anamaria Ciocan[1], Cristina Costescu[1], Daniel Hernandez Garcia[9], Pablo Gómez Esteban[7], James Kennedy[10#], Honghai Liu[11], Silviu Matu[1], Alexandre Mazel[12, 3], Mihaela Selescu[1], Emmanuel Senft[14], Serge Thill[15], Bram Vanderborght[7], David Vernon[16], Tom Ziemke[17]*,

Affiliations:

[1]Department of Clinical Psychology and Psychotherapy, Babeș-Bolyai University; Cluj-Napoca, Romania.

[2]Icahn School of Medicine at Mount Sinai; New York, USA

[3]School of Computer Science, University of Lincoln; Lincoln, UK.

[4]IDLab-AIRO, Ghent University & imec; Ghent, Belgium.

[5]School of Informatics, University of Skövde; Skövde, Sweden.

[6]Department of Computer Science, Loughborough University; Loughborough, UK.

[7]Brubotics, Vrije Universiteit Brussel & imec; Brussels, Belgium.

[8]College of Engineering, Can Tho University, Vietnam,

[9]School of Mathematical and Computer Sciences, Heriot-Watt University; Edinburgh, UK.

[10]Centre for Robotics and Neural Systems, Plymouth University; Plymouth, UK

[11]School of Computing, University of Portsmouth; Portsmouth, UK

[12]human-machine-interaction.org; Paris, France.

[13]Softbank Robotics; Paris, France.

[14]IDIAP Research Institute; Martigny, Switzerland.

[15]Donders Institute for Brain, Cognition & Behaviour, Radboud University; Nijmegen, Netherlands.

[16]Carnegie Mellon University Africa, Kigali, Rwanda.

[17]Department of Computer & Information Science, Linköping University; Linköping, Sweden.

Currently at Apple Inc., Pittsburgh, PA, USA.

* Corresponding author. Email: tom.ziemke@liu.se.

Abstract: The use of social robots in therapy for children with autism has been explored for more than 20 years, but there still is limited clinical evidence. The work presented here provides a systematic approach to evaluating both efficacy and effectiveness, bridging the gap between theory and practice by targeting joint attention, imitation and turn-taking as core developmental mechanisms that can make a difference in autism interventions. We present two randomized clinical trials with different robot-assisted therapy implementations aimed at young children. The first is an efficacy trial (N = 69; mean age = 4.4 years) showing that 12 bi-weekly sessions of in-clinic robot-assisted therapy achieve equivalent outcomes to conventional treatment, but with a significant increase in the patients' engagement. The second trial (N = 63; mean age = 5.9 years) evaluates the effectiveness in real-world settings by substituting the clinical setup with a simpler one for use in schools or homes. Over the course of a modest dosage of five sessions we show equivalent outcomes to standard treatment. Both efficacy and effectiveness trials lend further credibility to the beneficial role that social robots can play in autism therapy, while also highlighting the potential advantages of portable and cost-effective setups.

One-Sentence Summary: Robot-assisted therapy for children with autism is shown to be equally efficacious and more engaging than conventional therapy.

Main Text:

INTRODUCTION

This paper reports on two large-scale clinical trials comparing robot-assisted therapy (RAT) for children with autism spectrum disorder (ASD) with current standard psychological interventions, in both formal (clinical) and informal (educational) settings. The use of the robots in the treatment of ASD has been investigated by researchers for about 25 years (*1-5*), and the general idea is both evolutionary and challenging. Considering autistic children's general interest in technology, and the fact that interaction with social robots is considered safe and enjoyable (*6, 7*), it has been argued by several authors that the involvement of social robots in psychological interventions can lead to better results and cost-effective interventions in achieving therapeutic and educational objectives (*8, 9, 5*). However, it is challenging to include social robots in the therapeutical process, since good collaboration between end-users, psychologists, teachers and researchers is required to design and build appropriate and beneficial interventions for children with autism. The general idea of RAT has instilled high hopes, not least in parents, for improving the current services for children with autism. However, when we analyze the effectiveness of social robots in autism therapy reported in randomized clinical trials, the results are mixed, meaning that both positive and negative outcomes have been reported (for recent reviews see *10, 5, 11*). Although there have been some trials with positive results, there is certainly a need for trials that better integrate robot-assisted activities in the therapeutical process, explore the core developmental mechanisms, integrate them in educational contexts, and explore long term results (*12*). Without sufficient data, there is a risk of instilling false hopes and causing unjustified costs for patients and their families. A recent meta-analysis on the use of social robots for autistic children (*5*) suggests that more trials are needed to identify specific characteristics of RAT (e.g., type of robots, clear and transparent sample selection, settings) as well as better integration techniques in the therapeutic process. This paper contributes to filling that gap.

Most previous studies on RAT for children with ASD are proof-of-concept studies, clinical cases/single-case experiments, exploratory studies with small samples, experimental rather than clinical interventions, and often not guided by the evidence-based psychotherapy framework targeting established key mechanisms, such as joint attention, attention, and turn taking (*13*). Moreover, the interventions investigated in these studies target a very broad range of different abilities (e.g., play skills, social interaction abilities, engagement) which are typically defined differently in different studies, leading to inconsistent results that are not easy to integrate. These inconsistencies and the lack of standardized instruments in measuring the outcomes, constitutes a substantial challenge for practitioners trying to assess the applicability and effectiveness of RAT in specific cases. Although there are several studies with standardized protocols, large samples of children, and standardized measurements (e.g., *12, 14, 15*), there is still a shortage of studies rigorously comparing RAT to conventional therapy for children with ASD (in the following, we refer to the latter as 'standard human therapy' or SHT hereafter; cf. below). Moreover, as a recent systematic review pointed out, "while significant efforts have been made to develop and test the technical aspects, clinical validation and implementation lag behind" (*11*). A recent study by Ghiglino and colleagues tested the effects of robot-assisted activities focusing on visuospatial perspective-taking, integrated into the

rehabilitation program offered in a healthcare institution, and their results indicate that children with ASD gain the most benefit from activities incorporated into interactions with a humanoid robot, compared to standard therapy or other interventions.

Despite the differences observed between conventional treatment and RAT in some previous studies, the clinical meaningfulness of these findings is limited in the sense that the core developmental mechanisms of change still remain unknown (*14*). Hence, there is still a lot of uncertainty regarding how RAT really compares to SHT. We addressed some of the limitations in this domain in a large European project called DREAM ("Development of Robot-Enhanced therapy for children with AutisM spectrum disorders"), through both clinical and technological innovations (see also *9, 16*), which have been integrated and evaluated in the two clinical studies reported in this paper. One of the strengths of our work presented here, we believe, lies in the conceptualization of RAT in accordance with the evidence-based psychotherapy paradigm in clinical psychology, meaning that we used both clinical and psychometric methods for a rigorous clinical assessment of ASD. Previous studies focused either on clinical symptoms of ASD or only basic skills. Furthermore, some recent studies tested the effectiveness of social robots in home-based therapy, an environment which is slightly different from our approach (*17,18*). We fully support the general idea of using social robots in home environments in the future, but we also strongly believe that at this point more research is needed to investigate the underlying mechanisms of RAT to be used in home settings, where other types of challenges can appear (*19)*. Central to our work is a clear distinction between the clinical symptoms of ASD (as documented in DSM-5, the Diagnostic and Statistical Manual of Mental Disorders, fifth edition) and the underlying mechanisms. Our hypothesis, based on previous work (e.g., *20*), is that imitation, joint attention, and turn-taking are crucial fundamental abilities to train in children with ASD. However, no previous study investigating the use of social robots has treated these abilities as the main targets. These fundamental abilities, we believe, also constitute a base for developing more complex skills, such as play and social interactions, and they contribute to the cognitive and social development of children and predict their later communication and intellectual functioning.

Some previous studies have investigated the robot-mediated response to joint attention in a small sample (20) of young children with ASD (21), with small but not significant group differences being observed when using a Nao robot to elicit joint attention responses. The study took place in a clinical setting and all participants underwent four sessions of 10 minutes each. Two recent studies implemented by the same research group investigated the effectiveness of pivotal response treatment with or without robot assistance for improving social skills in children with ASD (*12*, *22*). One of the studies investigated the effectiveness of robot-assisted pivotal response treatment compared to pivotal response treatment in improving general social-communication skills and self-initiations measured with both standardized questionnaires and observational grids (*22*). Fifty-two children were assigned to one of the groups and participated in 20 intervention sessions in which self-initiations were targeted using motivational techniques from pivotal response treatment. Their results showed no significant differences between the two groups, although both interventions were effective in improving children's general socio-communication skills. No significant associations were found between the self-initiations and overall communication skills. Therefore, future studies are needed to further investigate the prerequisites and underlying mechanics of the social deficits in ASD. Similar results were identified when comparing robot-assisted pivotal response treatment

with treatment as usual (*12*) in a study involving 73 children with ASD. Both studies were conducted in clinical settings and limited their intervention protocols to self-initiations (e.g., asking for objects, asking for help by using utterances).

Another important issue that we try to address in our studies is the generalizability of the results, since our final goal is to enhance social skills of autistic children in real-world human-human social interactions (*3, 23*). Considering that the focus of our intervention is on hypothesized underlying mechanisms (imitation, joint attention, turn-taking), which support social skills acquisition (i.e., learning directly various social/communication behaviors), rather than on clinical symptoms as such, we are conceptualizing the intervention as targeting core developmental mechanisms rather than symptoms. We believe that this can support flexibility and generalizability in learning new social communication behaviors. Consequently, the intervention here is not conceptualized as a series of experimental sessions, but as a minimum clinical sequence, which could then be generalized to a full long-term or intense treatment. On that account, we expect to identify differences in the clinical symptoms, even though our approach is not a direct and long-term one. Another strength of our approach is that we compare RAT and SHT, both in an efficacy study (i.e., in well-controlled lab conditions) and in an effectiveness study (in a school setting, i.e., in more ecological, real-world conditions), in line with the most rigorous conditions of evidence-based therapy. The work presented here, documenting two complementary clinical studies, addressing efficacy and effectiveness, respectively, contributes to the integration of both these needs: a highly rigorous methodology and effectiveness in real-world settings.

Furthermore, considering the approach to target core developmental mechanisms, we aim to clarify the active elements from the interventions that could help children with autism improve their abilities and reduce their social difficulties. We propose an approach that allows us to identify and confirm the hypothesized underlying mechanisms (imitation, joint attention, and turn-taking) of change in a controlled efficacy study and subsequently to maximize those active elements in an effectiveness study. This approach allows us to clarify what abilities are most important to train in RAT.

In this process, the robot is used as a mediator. As previously mentioned, we strongly believe that if social robots are to be part of the therapeutic process, they need to be adopted by practitioners in their daily activities and enable them to actively participate in the design of the robot-assisted activities (*24*). Importantly, the mechanisms and skills learned in interaction with the robot should be generalizable to human-human interactions with the help of the therapist and should not remain limited to human-robot interactions only.

In the development of our studies, we followed a participatory design with the following steps: first, the initial concepts were developed by the psychologists and engineers, after which feedback from teachers, parents and children was gathered. The tasks in pilot studies (see *25, 26*) were then tested, refining the prototypes based on the feedback and results from these pilot studies. Finally, the intervention program was adapted and scaled to include more children.

RESULTS

Our approach moved from a Wizard of Oz paradigm, in which the robot is controlled by a human operator, to a semi-autonomous one, in which the robot's autonomous decision-making is supervised by a human therapist. Our platform for delivering semi-autonomous RAT for children with ASD has been made available as open-source and a reusable platform (*16*).

More specifically, to carry out the studies reported here, we developed two technical platforms (see Fig. 1): a stationary system used in controlled clinical trials to evaluate the efficacy of RAT (hereafter referred to as the clinical trials platform), and a smaller, more portable system that is used to evaluate the effectiveness of RAT in real-life educational settings (the educational platform). The clinical trials platform (*25,9, 26, 16*) consists of a sensorized table (Fig. 1A) on which the robot is placed, equipped with several cameras that record aspects such as the pose, movement and facial expressions of the child. This sensory information is used by the system to assess whether the behavior of the child is in accordance with the expectations of the clinical intervention scripts. That information in turn is used to automatically decide the next actions of the robot, thereby enabling the robot's semi-autonomous behavior. However, therapists always retain the ability to override the system's decisions. The educational platform (*27*), on the other hand, omits the sensorized table, and the robot is instead controlled via a tablet (Fig. 1B). This makes the educational platform more portable, but all clinically relevant attributes are retained, as discussed in more detail below.

Fig. 1. Robot-assisted therapy platforms. (**A**) Clinical trials platform used in Study 1. (**B**) educational platform used in Study 2.

Study 1: Efficacy of Robot-Assisted Therapy

Intent-to-treat analysis on all outcomes indicated that the two interventions (RAT vs. SHT) are equivalent on all primary outcomes and on clinical outcomes. However, the RAT intervention was superior on several secondary outcomes, reflecting engagement with the intervention (see Fig. 2). Statistical analysis and results are described in more detail in the supplementary material.

SHT was structured as a behavioral teaching program rooted in applied behavior analysis (ABA) discrete trial training methods, an approach well-established in autism interventions in Romania. The treatment directly targeted core social-cognitive domains—imitation, joint attention, and turn-taking—ensuring alignment with the robot-assisted therapy condition. Sessions consisted of structured play tasks facilitated by a human therapist, where learning followed the discriminative stimulus → child response → contingent reinforcement format. Correct responses were reinforced through positive social feedback (e.g., praise, smiles, gestures), whereas incorrect or no responses were supported through shaping, modeling, and prompting. These skills were practiced through repetitive, reinforcement-based exercises embedded in natural play contexts to promote consolidation, motivation, and engagement. Importantly, the structure and dosage of sessions (12 in total, with evaluation, training, and post-evaluation phases; each lasting 45 minutes) were intentionally matched to the RAT condition, ensuring comparability in treatment intensity and exposure.

Regarding imitation, our results revealed that children from the RAT group significantly improved their imitation ability compared to the SHT group ($F(1, 84.57) = 23.44$, $p < 0.001$). Also, the results showed that they significantly improved their performance from

initial assessment to final assessment ($F(9, 110.95) = 5.38$, $p < 0.001$), and they improved their performance with the increase in difficulty from level 1 to level 2 ($F(2, 84.81) = 4.03$, $p = 0.021$). Comparisons for the group and time interaction indicate that the scores for the participants in the SHT group did not change from baseline assessment during the intervention, while participants in the RAT group reported significant improvements starting with session 3 ($p = 0.004$), and performance remained significantly higher until the final assessment ($p < 0.001$). Comparisons conducted between groups, suggested that children in the RAT condition had significantly better performance starting with intervention session 2 ($p = 0.032$), and maintained this advantage until the end of the study ($p < 0.001$). In the case of joint attention, there were no significant differences between SHT and RAT groups. Comparisons inside the SHT group showed decreases in scores when comparing initial assessment with intervention session 5 ($p < 0.001$), session 7 ($p < 0.001$), session 8 ($p < 0.001$), and final assessment ($p = 0.003$). No within-group differences emerged for the RAT group when comparing intervention sessions and the final assessment with baseline scores. The RAT intervention had better scores for the two final intervention sessions compared to SHT; however, the difference was no longer apparent in the final assessment.

For turn-taking, both groups improved their performance during the interventions starting with the first half of the sessions. The overall performance was lower in the information sharing task (mean= 0.48, standard error = 0.03), compared to categorization (mean= 0.66, standard error = 0.3) and pattern completion (mean = 0.65, standard error = 0.30). The SHT group showed some advantages, especially in the pattern completion task, but there were no significant overall or final assessment differences between the two groups.

Fig. 2. Study 1 results: efficacy of robot-assisted therapy using the clinical trials platform – primary outcomes (social skills, clinical outcomes, satisfaction, and perceived effectiveness) and secondary outcomes (eye-contact and smiling, i.e., engagement)..

Regarding the results for the secondary outcomes (eye-contact and smiling) in the imitation tasks, children maintained eye-contact better in RAT sessions (mean = 8.01, standard error = 0.53) compared to SHT (mean = 5.54, standard error = 0.53). Similar results were obtained for smiling: children in the RAT group smiled more (mean = 7.72, standard error = 0.58), compared to children in the SHT group (mean = 3.80, standard error = 0.54). In the joint attention tasks, children also seemed to enjoy the RAT sessions more, showing increased levels of eye-contact (mean = 4.21, standard error = 0.33), compared to SHT sessions (mean = 2.94, standard error = 0.32), and they smiled more (RAT sessions mean = 4.25, SE = 0.34, in SHT sessions mean = 2.60, standard error 0.33). The same pattern emerged in the turn-taking tasks: even though there were some differences between the types of turn-taking tasks, overall children were more engaged in the RAT tasks showing more eye-contact episodes (mean = 12.12, standard error = 0.98) compared to SHT sessions (mean = 6.93, standard error = 0.96) and smiled in RAT sessions (RAT mean = 15.33, standard error = 1.27, SHT mean = 8.44, standard error = 1.24).

When investigating the differences in clinical symptoms, as illustrated in Fig. 3, we did not find any significant differences between SHT and RAT groups ($F(1, 60.75) = 3.15$, $p = 0.381$); however, our results revealed that the scores decreased from pre-test (mean = 14.23, standard error = 0.48) to post-test (mean = 12.60, standard error = 0.52) for both

groups. A similar pattern was observed when analyzing the SCQ results, scores decreased from pre-test (mean = 16.20, standard error = 0.70) to post-test (mean = 14.56, standard error = 0.81). Both the scores from ADOS and from SCQ are very close to the cut-off scores (for autism diagnoses) reported in some studies.

Fig. 3. Study 1 results: efficacy of robot-assisted therapy using the clinical trials platform – clinical outcomes (autism disorder observation schedule ADOS and social communication questionnaire SCQ).

In summary, the results indicate that RAT is as effective as SHT in changing some of the symptoms and underlying mechanisms (joint attention, imitation, turn taking) of ASD. This is important for both RAT and SHT, taking into account the low psychological treatment (two initial evaluation sessions, eight intervention sessions and two final evaluations sessions). RAT is more effective than SHT in improving various engagement mechanisms. Considering that treatment protocols for ASD are intense or long, our results suggest using RAT independently or as an additional alternative to SHT (e.g., when there is a decrease in motivation during SHT).

Study 2: Effectiveness of Robot-Assisted Therapy

Having established the efficacy for some mechanisms in RAT in the laboratory in Study 1, we turned to investigating stability and generalizability in real-life settings in Study 2. We conducted an effectiveness study implemented in 10 special education institutions and therapy centers for ASD children in Cluj-Napoca, Romania. The children were randomly allocated to one of two conditions: RAT intervention delivered via the educational platform (including treatment as usual) and a control group receiving only treatment as usual. All children were recruited from special education institutions, meaning that they all followed regular therapeutic and educational programs outside of the study protocol. Participating children had a history of diagnosed ASD symptoms and were able to understand the instructions required for assessment as well as for delivering the RAT intervention. Therefore, in contrast with Study 1, here we tested RAT in real-life conditions, for a larger age range, and in a transdiagnostic framework (i.e., symptoms of ASD, often comorbid with other clinical conditions). Moreover, this study was conducted as a superiority study, and we expected that the addition of RAT to the standard treatment as usual intervention would improve outcomes.

Treatment as usual in the special schools from Romania, where data was collected, entails a structured approach that combines individualized educational programming with therapeutic and social support services. Typically, children benefit from daily participation in small-group and one-on-one instructional activities led by specialized teachers and therapists. The curriculum is adapted to each child's developmental level, focusing on core educational domains (such as communication, social skills, self-care, and academic abilities) and integrating established intervention models like ABA, speech and language therapy, occupational therapy, and social skills training. Intervention strategies may include structured teaching, visual supports, repetition, reinforcement-based learning, and the use of clear routines and predictable schedules to enhance adaptation and learning. The intervention environment typically emphasizes low student-teacher ratios, individualized supports, and collaboration between educators, therapists, and families to maximize developmental gains and social inclusion.

The results indicated that the RAT group outperformed the control group on imitation abilities. No significant differences between the two groups appeared for joint attention and turn-taking. Results regarding the hypothesized underlying mechanisms are presented in Figure 4. Results regarding social interaction difficulties, as rated by parents and teachers, are provided in Figure 5.

Fig. 4. Study 2 results: effectiveness of robot-assisted therapy using the educational platform – performance in imitation, joint attention, turn-taking, and emotion recognition.

Fig. 5. Study 2 results: effectiveness of robot-assisted therapy using the educational platform – Performance in social interaction assessed by parents and teachers.

Children improved their imitating ability from initial evaluation to final evaluation in the RAT group (pre-test mean = 2.18, standard error = 0.10 to post-test mean = 2.45, standard error = 0.10). However, there was no significant difference between the RAT group and the treatment as usual group ($F(1, 61.44) = 0.64$, $p = 0.427$). Similar results were obtained for turn-taking skills, meaning that there were significant improvements from initial to final assessment (pre-test mean = 2.20, standard error = 0.12 to post-test mean = 2.53, standard error = 0.12), but no significant differences between the treatment as usual group and RAT group ($F(1, 60.77) = 0.07$, $p = 0.789$). For joint attention there were no significant differences from initial to final assessment ($F(1, 59.64) = 3.20$, $p = 0.079$) individually or between groups ($F(1, 60.91) = 0.01$, $p = 0.933$). Regarding emotion recognition, the results indicated no significant effects: for time ($\mathbf{F}(1, 60.15) = 0.83$, $p = 0.367$), group ($\mathbf{F}(1, 62.79) = 0.15$, $p = 0.703$), and their interaction ($F(1, 60.15) = 0.449$, $p = 0.506$).

The results revealed a significant effect for time, both interaction skill perceived by parents ($F(1, 51.52) = 4.88$, $p = 0.032$) and by teachers ($F(1, 53.86) = 12.14$, $p < 0.001$), but no significant effect between the treatment as usual control group and RAT, either for parents ($F(1, 57.17) = 3.60$, $p = 0.063$) or for teachers ($F(1, 57.67) = 0.001$, $p = 0.971$).

We found in Study 2 that RAT was well accepted among children, parents, teachers, and therapists. Both RAT and treatment as usual led to improvements from pre- to post-test, on some of the mechanism (imitation and turn-taking), and in the evaluation of social skills by both parents and teacher. This is an important finding considering the low psychological treatment dosage in Study 2 (one initial assessment, three interventions, and one final evaluation session). RAT had a greater effect than the control group (TAU) on imitation.

DISCUSSION

The two studies aimed to investigate the efficacy and effectiveness of RAT for children with autism compared to conventional standard human therapy and treatment as usual. In this paper, we have proposed and evaluated an approach that bridges the gap between theory and practice, with a focus on training the underlying mechanisms (imitation, joint attention, and turn-taking) that can make a difference in autism interventions. This work validated a standardized protocol for RAT for children with autism that can improve their imitation skills. Although some previous studies demonstrated the effectiveness of social robots in developing joint attention (*39-40*), we did not find any significant difference compared to standard human therapy in our work. One important finding from

investigating joint attention performance during the RAT sessions in comparison with standard human therapy sessions is that in the latter children tend to lose interest and can even perform worse after several sessions than in the initial evaluation due to the repetitive tasks, whereas with RAT they maintain their interest. Their increased interest can be observed in their performance in the joint attention tasks as well as their increased positive emotions (e.g., smiling) and prolonged eye-contact. One possible explanation is that our intervention protocol used standardized tasks, and recent studies have shown the need to personalize the intervention according to children's needs (*41, 15*) in order to obtain a better influence on children's abilities. A recent study (*12*) reported similar results as ours: no significant differences between standard human therapy and RAT in a large sample after treatment measures. One possible explanation is the fact that we included intensive interventions in our treatment as usual conditions, which is different from previous studies, where control groups (*42*) or community-based or ABA-focused interventions (*43, 44*) were used.

Another important finding pertains to the measured decrease in clinical symptoms of ASD following RAT in Study 1. The results highlight significant improvements in core developmental mechanism-level skills—imitation, joint attention, and turn-taking—which served as our primary outcomes. However, generalization to broader, functional domains was not directly assessed. Although reductions in clinical symptom scores on ADOS and SCQ approached relevant cut-offs, suggesting promising potential, improvements observed in trained tasks do not necessarily equate to generalized social or clinical gains, as recent evidence and reviewer feedback indicate. The lack of group-level differences on validated distal clinical scales further suggests that RAT's effects in this study primarily reflect task-specific learning rather than broad clinical change. We acknowledge this limitation and concur with the need for future studies to incorporate independent generalization measures and assessments of real-world social functioning in diverse, naturalistic environments. Nonetheless, our approach—focused on repeated engagement and systematic training of core social-cognitive skills—may offer an innovative pathway for targeted skill acquisition and technology-based intervention development for children with autism. These findings invite further research into how the presence of robotized supports, even in school settings and for short periods, can facilitate foundational social learning, which will require additional validation through generalization-focused, ecologically meaningful outcome measures. Overall, our results indicate the effectiveness of our approach to RAT both in controlled and real-life conditions.

As previously seen in other studies (*45*), RAT outperforms SHT in controlled conditions (Study 1) for various engagement measures, which is important from a long-term perspective, to keep children engaged. Also, the engagement and the novelty effect were thoroughly investigated in previous studies in the domain of RAT, and our results reveal that, even if the tasks are repetitive, children were still motivated and interested to continue in the RAT sessions. In real-world conditions (Study 2) RAT was found to outperform SHT (in the form of treatment as usual) in case of imitation. This result can contribute to future development of RAT interventions, suggesting that some of the underlying mechanisms are more appropriate to be trained using social robots.

One important limitation of our work is the absence of a no-treatment control group. Therefore, our results should be interpreted with caution, since the outcomes could be also influenced by other factors. Future studies should consider including a treatment as usual group (which we did not include in our Study 1) as well as a no-treatment control group.

Despite this limitation, the findings provide valuable insights and serve as a foundational basis for further exploration into the use of robot-assisted intervention for children with neurodevelopmental disorders. Moreover, children from our treatment as usual group received evidence-based interventions, so when compared to evidence-based interventions even a similar performance or a small effect on the RAT group represents an important and relevant gain.

The fact that the training program used here specifically targeted mechanisms such as imitation, joint attention, and turn-taking, subsequently used as metrics to evaluate improvement, could also be considered a drawback. This overlap between the intervention focus and the assessment metrics may introduce bias, potentially influencing the perceived effectiveness of the training. This limitation underscores the need for future research to incorporate a broader range of outcome measures, including those that assess generalized and functional improvements beyond the directly targeted skills. The measurements used in this study for clinical symptoms were ADOS and SCQ, instruments that were not so sensitive to the improvements made by children in RAT sessions in order to show the difference between the two groups.

Another limitation of our studies – as for most other studies – is related to the reduced dose of the treatment (intensity or duration). In future work, these protocols should be tested in more intense clinical interventions for longer durations. On the other hand, this can also be considered a strength, because it has allowed us to test both our RAT approach's internal validity (in the efficacy study) and the external validity (in the effectiveness study).

A significant limitation of Study 2 is the small number of intervention sessions provided. With only three intervention sessions, it is challenging to determine the long-term effectiveness and sustainability of the observed improvements. The short duration of the intervention may not have been sufficient to evaluate lasting changes or fully capture the potential benefits of the training. Consequently, the results might reflect initial gains rather than enduring improvements in the targeted outcomes.

A final limitation of the work presented here lies in the standardized approach used for all participants, as mentioned above. Recent studies highlight the importance of tailored intervention, and future studies should consider the co-design of the intervention session together with end-users (*46, 47*). Effective treatment for ASD often requires personalized interventions tailored to individual needs, especially because ASD encompasses a wide range of characteristics and varies significantly from individual to individual. Rigid protocols may fail to provide the personalized attention necessary for effective intervention, potentially leading to suboptimal outcomes.

A critical consideration in evaluating the effectiveness of RAT is determining whether the observed increase in children's engagement is due to the unique features of the robot itself or simply the novelty of the activity. In our study it is unlikely to be a novelty effect because we see an increase in engagement over several sessions. Rather, the robot's design and interactive capabilities, and the way the tasks are built, appear to intrinsically motivate and sustain engagement. This suggests a more durable and substantive therapeutic benefit, demonstrated by children showing increased engagement over several intervention sessions.

Overall, the results contribute to an understanding of the relevance of technological solutions in treating children with ASD and reducing the human workload that traditional therapies impose. Evidence-based psychological treatments for ASD (e.g., derived from applied behavioral analysis) are intense and of long duration. Many ASD patients and their families do not have access to necessary treatments, cannot afford the costs, do not respond well, or relapse. In the absence of available therapy, many children diagnosed with ASD are treated solely with medication, despite the lack of persistent effects. In summary, the work presented here constitutes one of the largest studies on RAT in autism treatment, further extending the existing clinical evidence base, and it provides a systematic approach to evaluating both efficacy and effectiveness in an integrated manner through the combination of complementary clinical studies using a unified approach. Future studies could consider more than the three skills addressed here (or others) in their protocols, and there is still scope to personalize the interventions according to children's needs.

MATERIALS AND METHODS

Study 1 Efficacy of Robot-Assisted Therapy using the Clinical Trials Platform

In Study 1, using the clinical trials platform, we investigated the efficacy of RAT compared to SHT in an equivalence trial design. We established an equivalence margin of Cohen's $d = 0.79$, meaning that the maximum standardized difference between the two groups should not be larger than a medium effect size. The clinical trial protocol was registered on ClinicalTrials.gov with the number NCT03323931. The clinical study received prior ethical approval from the Scientific Council of Babes-Bolyai University in Cluj-Napoca, Romania, where the trial was conducted (record no. 30664/February 10th, 2017). A letter of consent was signed by at least one parent before the study, expressing their consent to record assessments and interventions and to use the data and recordings for scientific purposes in an anonymous fashion.

Sample

Participants were between 2.5 and 7 years and had a formal diagnosis of autism in their medical record, confirmed by an assessment with the Autism Disorder Observation Schedule (ADOS) (28) carried out by clinical psychologists with an ADOS certification (and blind to the study design). The majority of the children had a score of 12 or above (51 children) meaning that they have a moderate-to-substantial need of support (*29*). Eighteen children had scores from 7 to 11, meaning that they only need minimal support from the caregivers. Children were excluded if they had other comorbid neurodevelopmental or learning disorders, if they had high levels of social skills at the initial assessment, as indicated by perfect performance on all the assessment tasks, or if they lacked the ability to understand the task requirements and follow the instructions in a session of about 10-15 minutes. Figure 6 presents a CONSORT diagram providing details on the study participants. The final sample included 69 participants, of which 12 were female (17.4%), with a mean age of 52.58 months (standard deviation = 11.66).

Fig. 6. Study 1: Efficacy of robot-assisted therapy. CONSORT Diagram showing how participants were enrolled, allocated to the SHT and RAT groups, excluded or discontinued, and analyzed.

Measures

Primary outcomes were measured in terms of imitation performance, joint attention performance, and turn-taking performance.

Imitation performance was measured using a dichotomic scale (correct or incorrect answer). A good performance rating was achieved when the child had correctly imitated the behavior of the play partner (human or robot), including both the movement and the sound made by the partner. For each session, an average score was calculated for each child and used in data analysis The average score is obtained by dividing the number of correct behavioral imitations by the number of attempts.

Joint attention performance was measured on the same dichotomic scale. Children received a positive rating if they followed the cues shown by the interaction partner. A correct answer counted if the child looked at an object displayed on the left or the right side of large touchscreen placed in front of them. The eye-tracking algorithms of the clinical trials platform were used to detect the eye-gaze of the child (*30*), whereas the observing clinician used visual input coming from the left and right cameras (positioned with respect to the child), to decide on the rating of the behavior. Similar to imitation, an average score was calculated for each child in each session.

Turn-taking performance was also measured as correct or incorrect for each trial with the help of the sensorized table, supervised by the clinician. Children received a positive performance rating if they waited for the partner's turn in the game without touching the screen, while the partner performed the action. The semi-autonomous supervised system determined whether the child waited their turn by means of skeleton and posture analysis, based on the video and depth cameras in the setup (*30, 31, 32*). An average score for each child for each session was also used for turn-taking in data analysis.

Secondary outcomes included the engagement of the child, measured in terms of eye-gaze and smiling at the play partner during the sessions. Engagement measures were chosen so that they offer an estimate of the degree to which the child is focusing on the treatment activities and is enjoying them, rather than being distracted, getting involved in other activities or feeling distressed during the intervention. Eye-contact with the play partner was measured in real time (*30*). An eye-contact event was defined by the eye-gaze of the child directed towards the head of the play partner for at least 200 milliseconds. The presence or absence of an eye-contact event was determined using eye-tracking algorithms. Smiling during sessions was also considered an indicator of engagement and was measured in real time by the sensorized table, detecting smiling expressions based on the facial landmarks collected from the cameras during the intervention (*33*).

The Autism Disorder Observation ScheduleADOS (*28*) is a set of standardized activities designed to cue behaviors that are relevant for the diagnosis of ASD, related to social behavior and communication (e.g., responding when one's name is called, asking for help to get access to food, responding to social smiles, etc.). ADOS comprises four modules which have different activities adapted to the age and the level of verbal abilities of the child. In the current study, only Module 1 (for children who do not use phrase speech) and Module 2 (for children who make spontaneous use of meaningful short phrases of up to three words) were used. ADOS offers a coding and rating system tied to clinical criteria for the diagnosis of ASD, allowing the clinician to rate the level of severity and establish a

likely diagnosis, by comparing the score of the child with the cut-off scores established by research.

The Social Communication Questionnaire SCQ (*34*) is a questionnaire consisting of 40 items on which a parent of the child rates the presence and the absence of normal and abnormal behaviors which are relevant for ASD. Two forms are available: "Current*"* and "Lifetime". The former looks at the child's last 3 months, whereas the latter looks at the child's entire developmental history. The "Current" symptoms version of the questionnaire which we used to assess parents' perceptions over the changes in ASD symptoms has shown poorer diagnostic utility in children under 4 years (*35*), and thus we did not use its cut-off scores. Instead, following other studies (*36*), we used this scale as a continuous measure of symptoms as reported by parents to gather information about the generalizability of the intervention effects beyond the interaction with the therapist.

To determine parents' perceptions of child improvements and satisfaction, at the end of the intervention we asked one parent of each child to rate the perceived improvements of the child overall (one item), and on the three specific skills (three items), using a self-report scale from 1 ("I didn't see any progress") to 5 ("The progress was very good"). We also asked parents to rate their satisfaction with the treatment that the child received using 10 items, each asking about satisfaction with a specific component (e.g., how assessment was conducted, the assessment reports they received, the degree to which the intervention was tailored for the needs and skills of the child). The satisfaction for each item was rated on a scale from 1 (not at all satisfied) to 5 (very satisfied). Both scales were adapted for the current study and showed excellent reliability, with Cronbach's alphas of 0.88 for perceived improvements and 0.93 for satisfaction.

Procedure

All recruited participants were required to attend 12 sessions: 2 initial evaluation sessions (i.e., pre-intervention), 8 intervention sessions, and 2 final evaluation sessions (i.e., post-intervention). The 45 minutes sessions were held bi-weekly. The overall structure of the 12 sessions was similar for both groups. Sessions 1 and 2 were designed to carry out a comprehensive psychological evaluation and to determine the baseline level of the three fundamental social abilities (imitation, joint attention, and turn-taking). ADOS was administrated to each child to assess the social and communicative behavior associated with ASD. In addition, the SCQ was completed by the children's parents/caregivers. Sessions 3 to 10 were used for training imitation, joint attention, and turn-taking. Sessions 11 and 12 first determined the level of the three abilities after the interventions through tasks in which the targeted behaviors were preceded by a stimulus but not followed by any feedback. In the post-intervention assessment, ADOS and SCQ were used to determine if the benefits of the interventions were generalized.

Intervention sessions with the robot assistant had a similar structure and followed the principles of discrete trial training (*37*). Intervention tasks were introduced as gameplay activities. The interaction partner presented a discriminative stimulus and waited for the child's response. If the child response matched the expected one, then the robot provided positive social feedback (e.g., praise). If the child's performance was below the expected one, then the robot provided encouragement to try again. The child was given three opportunities to respond to each exercise. After the third failure, the therapist guiding the intervention prompted the child's behavior.

Study 2: Effectiveness of Robot-Assisted Therapy using the educational platform

Sample

The study sample included 63 participants, of which 7 were females (11.11%), with a mean age of 5.89 years (standard deviation = 2.63). The sample size was calculated using *G**Power with the following specifications: $\alpha = 0.05$, 1-$\beta = 0.80$, and an estimated effect size for the population of at least Cohen's $d = 0.80$. Children were included if their age was between 3 and 12 years, if they had a history of diagnosed ASD symptoms, and were able to understand the instructions required for assessment and for delivering the RAT intervention. Fig. 7 presents the CONSORT Flow Diagram providing details on the participants.

Fig. 7. Study 2: effectiveness of robot-assisted therapy using the educational platform. CONSORT Flow Diagram showing how participants were enrolled, allocated to the treatment as usual group and the RAT group, excluded or discontinued, and analysed.

Measures

The primary outcomes were constituted by the same basic underlying mechanisms as in Study 1 (imitation, joint attention, and turn-taking), measured with specific tasks derived from ADOS. Moreover, we added an additional emotion recognition measure related to the recognition of emotional expressions. As secondary outcomes we measured the difference between initial assessment and final assessment of the clinical symptoms of social interaction, as rated by parents and teachers.

Emotion recognition was measured during the interaction sessions when children had to indicate the emotions expressed by the robotic play partner - happy, sad, scared, or angry. Children who were verbal could do so by naming the emotion, whereas non-verbal children were given a set of photocards to express each of the four possible emotions. -

To assess the interaction skills as perceived by parents and teachers, for each child in each group, we asked one parent and one teacher to assess the child's ability to interact with others using a set of items extracted from the TRIAD Social Skills Assessment that we used for parents and teachers (*38*; and specifically, we used items 44-49 from the parent's module and items 45-50 from the teacher's module). The items were identical for both raters, and we asked them to indicate how competent the child is, on a scale from 1 (not very competent) to 4 (very competent), in expressing feeling, understanding the perspective of others, initiating social interaction, among others. Both raters filled out the items before and after the intervention. The scale was translated and adapted into Romanian language for the present study. The internal consistency, a measure of reliability that evaluates how closely related the items within a questionnaire are, was very good, with Cronbach's alpha above 0.80 for both parents and teachers at both points in time.

Procedure

The intervention was a shortened and simplified version of the one implemented in study 1. We started with a behavioral assessment of imitation, joint attention, turn-taking, and emotion recognition skills. After the initial assessment, we offered children three intervention sessions of approximately 30 minutes each, performed daily or every second day, which targeted these skills. This was followed by a final evaluation session identical to the initial one.

In Study 2 the robot was controlled by a research assistant present in the room using a DREAM Lite tablet application. The coding of the child's behavior (the degree to which the child imitated, initiated joint attention, and took turns) was made by a graduate level clinical psychologist, specifically trained for this purpose. All sessions were recorded with a video. Each team delivering the intervention, consisting of the therapist, the observer coding child's responses, and the research assistant controlling the robot, had weekly meetings with a supervisor, who was a PhD-level experienced clinical psychologist or a special education specialist. During supervision meetings, each individual case was discussed, and intervention sessions were reviewed based on the video.

Supplementary Materials (in separate file)

Materials and Methods

Supplementary Text

REFERENCES

1. Dautenhahn K (1999). Robots as social actors: Aurora and the case of autism. In *Proc. CT99, The Third International Cognitive Technology Conference*, August, San Francisco 1999 Aug (Vol. 359, p. 374).
2. Dautenhahn K & Werry I (2004). Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics & Cognition*, 12(1):1–35.
3. Diehl JJ, Schmitt LM, Villano M & Crowell CR (2012). The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Research in autism spectrum disorders*, 6(1):249–62. doi: 10.1016/j.rasd.2011.05.006
4. Scassellati B, Admoni H & Matarić M (2012). Robots for use in autism research. *Annual review of biomedical engineering*, 14:275–94. doi: 10.1146/annurev-bioeng-071811-150036.
5. Kouroupa A, Laws KR, Irvine K, Mengoni SE, Baird A & Sharma S (2022). The use of social robots with children and young people on the autism spectrum: A systematic review and meta-analysis. *PLoS One*, 17(6):e0269800. doi: 10.1371/journal.pone.0269800.
6. Clark M & Adams D (2020). The self-identified positive attributes and favourite activities of children on the autism spectrum. *Research in Autism Spectrum Disorders*, 72, 101512.
7. Jones M, Falkmer M, Milbourn B, Tan T, Bölte S & Girdler S (2022). The core elements of strength-based technology programs for youth on the autism spectrum: A systematic review of qualitative evidence. *Review Journal of Autism and Developmental Disorders*, 1-17.
8. Alabdulkareem A, Alhakbani N & Al-Nafjan AA (2022). Systematic Review of Research on Robot-Assisted Therapy for Children with Autism. *Sensors*, *22*, 944. https://doi.org/10.3390/s22030944.
9. Cao HL, Esteban PG, Bartlett M, Baxter P, Belpaeme T, Billing E, Cai H, Coeckelbergh M, Costescu C, David D, De Beir A, Hernandez D, Kennedy J, Liu H, Matu S, Mazel A, Pandey A, Richardson K, Senft E, Thill S, van de Perre G, Vanderborght B, Vernon D, Wakanuma K, Yu H, Zhou X & Ziemke T (2019). Robot-enhanced therapy: Development and validation of supervised autonomous robotic system for autism spectrum disorders therapy. *IEEE Robotics & Automation Magazine*, 26(2), 49-58.
10. Salimi Z, Jenabi E & Bashirian S. (2021) Are social robots ready yet to be used in care and therapy of autism spectrum disorder: A systematic review of randomized controlled trials. *Neuroscience & Biobehavioral Reviews*, 129:1–6. doi: 10.1016/j.neubiorev.2021.04.009.
11. Vagnetti R, Di Nuovo A, Mazza M & Valenti M (2024) Social Robots: A Promising Tool to Support People with Autism. A Systematic Review of Recent Research and Critical Analysis from the Clinical Perspective. *Rev J Autism Dev Disord* (2024). https://doi.org/10.1007/s40489-024-00434-5 (published online 29 February 2024).
12. van den Berk-Smeekens I, de Korte MWP, van Dongen-Boomsma M, Oosterling IJ, den Boer JC, Barakova EI, Lourens T, Glennon JC, Staal WG, Buitelaar JK (2022). Pivotal Response Treatment with and without robot-assistance for children with autism: a randomized controlled trial. *Eur Child Adolesc Psychiatry*, 31(12):1871-1883. doi: 10.1007/s00787-021-01804-8.

13. Robinson NL, Cottier TV & Kavanagh DJ (2019). Psychosocial health interventions by social robots: systematic review of randomized controlled trials. *Journal of Medical Internet Research*, 21(5), e13203.

14. Holeva V, Nikopoulou VA, Lytridis C, Bazinas C, Kechayas P, Sidiropoulos G, Papadopoulou M, Kerasidou MD, Karatsioras C, Geronikola N, Papakostas GA, Kaburlasos VG & Evangeliou A (2022). Effectiveness of a Robot-Assisted Psychological Intervention for Children with Autism Spectrum Disorder. *J Autism Dev Disord*, 1-17. https://doi.org/10.1007/s10803-022-05796-5.

15. Scassellati B, Boccanfuso L, Huang CM, Mademtzi M, Qin M, Salomons N, Ventola P & Shic F (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Science Robotics*, 3(21):eaat7544. doi: 10.1126/scirobotics.aat7544.

16. Billing E, Belpaeme T, Cai H, Cao H-L, Ciocan A, Costescu C, David D, Homewood R, Hernandez D, Esteban PG, Liu H, Nair V, Matu S, Mazel A, Selescu M, Senft E, Thill S, Vanderborght B, Vernon D & Ziemke T. (2020). The DREAM Dataset: Supporting a data-driven study of autism spectrum disorder and robot enhanced therapy. *PLoS ONE* 15(8): e0236939. https://doi.org/10.1371/journal.pone.0236939.

17. Clabaugh C, Mahajan K, Jain S, Pakkar R, Becerra D, Shi Z, Deng E, Lee R, Ragusa G & Matarić M (2019). Long-Term Personalization of an In-Home Socially Assistive Robot for Children With Autism Spectrum Disorders. *Front. Robot. AI* 6:110. doi: 10.3389/frobt.2019.00110.

18. Pakkar R, Clabaugh C, Lee R., Deng E & Matarić M (2019). Designing a Socially Assistive Robot for Long-Term In-Home Use for Children with Autism Spectrum Disorders. In: *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, New Delhi, India, 2019, pp. 1-7, doi: 10.1109/RO-MAN46459.2019.8956468.

19. Jain S, Thiagarajan B, Shi Z, Clabaugh C & Matarić MJ (2020). Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. *Sci. Robot*., 5,eaaz3791(2020).DOI:10.1126/scirobotics.aaz3791.

20. Poon KK, Watson LR, Baranek GT & Poe M. (2012) To what extent do joint attention, imitation, and object play behaviors in infancy predict later communication and intellectual functioning in ASD? *Journal of Autism and Developmental Disorders*, *42*, 1064-1074.

21. Zheng Z, Nie G, Swanson A, Weitlauf A, Warren Z, Sarkar N (2020). A Randomized Controlled Trial of an Intelligent Robotic Response to Joint Attention Intervention System. *J Autism Dev Disord*. 2020, 50(8):2819-2831. doi: 10.1007/s10803-020-04388-5.

22. de Korte MW, van den Berk-Smeekens I, van Dongen-Boomsma M, Oosterling IJ, den Boer JC, Barakova EI, ... & Staal WG (2020). Self-initiations in young children with autism during Pivotal Response Treatment with and without robot assistance. *Autism*, 24(8), 2117-2128.

23. Kumazaki H, Yoshikawa Y, Yoshimura Y, Ikeda T, Hasegawa C, Saito DN, Tomiyama S, An KM, Shimaya J, Ishiguro H, Matsumoto Y, Minabe Y & Kikuchi M (2018). The impact of robotic intervention on joint attention in children with autism spectrum disorders. *Molecular Autism* **9**, 46 (2018). https://doi.org/10.1186/s13229-018-0230-8.

24. Huijnen CAGJ, Verreussel-Willen HAMD, Lexis MAS & de Witte LP (2021). Robot KASPAR as Mediator in Making Contact with Children with Autism: A Pilot Study. *Int J of Soc Robotics* **13**, 237–249. https://doi.org/10.1007/s12369-020-00633-0.

25. David DO, Costescu CA, Matu S, Szentagotai A & Dobrean A (2018). Developing joint attention for children with autism in robot-enhanced therapy. *International Journal of Social Robotics, 10*(5), 595-605.

26. David DO, Costescu CA, Matu S, Szentagotai A & Dobrean A (2020). Effects of a robot-enhanced intervention for children with ASD on teaching turn-taking skills. *Journal of educational computing research, 58*(1), 29-62.

27. Mazel A & Matu S (2021). DREAM Lite: Simplifying Robot Assisted Therapy for ASD. *arXiv*:2104.08034v1, https://doi.org/10.48550/arXiv.2104.08034.

28. Lord C, Risi S, Lambrecht L, Cook EH Jr, Leventhal BL, DiLavore PC, Pickles A, Rutter M (2000). The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. *J Autism Dev Disord* **30**, 205–223. https://doi.org/10.1023/A:1005592401947.

29. Luyster R, Gotham K, Guthrie W, Coffing M, Petrak R, Pierce K, ... & Lord C. (2009). The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of autism and developmental disorders*, *39*, 1305-1320.

30. Cai H, Liu B, Zhang J, Chen S & Liu H (2015). Visual focus of attention estimation using eye center localization. *IEEE Systems Journal*, 11(3), 1320-1325.

31. Liu B, Ju Z, Kubota N & Liu H (2018). Online action recognition based on skeleton motion distribution. In *29th British Machine Vision Conference*. British Machine Vision Association.

32. Zhang S, Yu H, Dong J, Wang T, Qi L & Liu H (2015). Combining kinect and pnp for camera pose estimation. In *2015 8th International Conference on Human System Interaction* (pp. 357-361). IEEE.

33. Wang Y, Yu H, Stevens B & Liu H (2015). Dynamic facial expression recognition using local patch and lbp-top, *Proceedings of the 8th International Conference on Human System Interactions*, 362–367.

34. Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire*. Los Angeles, CA: Western Psychological Services.

35. Marvin AR, Marvin DJ, Lipkin PH & Law JK (2017). Analysis of Social Communication Questionnaire (SCQ) screening for children less than age 4. *Current developmental disorders reports*, *4*, 137-144.

36. Pickles A, Le Couteur A, Leadbitter K, Salomone E, Cole-Fletcher R, Tobin H, Gammer I, Lowry J, Vamvakas G, Byford S, Aldred C, Slonims V, McConachie H, Howlin P, Parr JR, Charman T, Green J (2016). Parent-mediated social communication therapy for young children with autism (PACT): long-term follow-up of a randomised controlled trial. *The Lancet*, *388*(10059), 2501-2509.

37. Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus on autism and other developmental disabilities*, 16(2), 86-92.

38. Stone WL, Ruble L, Coonrod E, Hepburn S, Pennington M, Burnette C, Brigham NB (2010). *TRIAD social skills assessment* (2nd ed). Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt Kennedy Center.

39. Kajopoulos J, Wong AHY, Yuen AWC, Dung TA, Kee TY & Wykowska A (2015). Robot-Assisted Training of Joint Attention Skills in Children Diagnosed with Autism. In:

Tapus, A., André, E., Martin, JC., Ferland, F., Ammi, M. (eds) Social Robotics. ICSR 2015. Lecture Notes in Computer Science(), vol 9388. Springer, Cham. https://doi.org/10.1007/978-3-319-25554-5_30.

40. Mehmood F, Ayaz Y, Ali S, De Cassia Amadeu R, & Sadia H (2019). Dominance in Visual Space of ASD Children Using Multi-Robot Joint Attention Integrated Distributed Imitation System. *IEEE Access*, 7, 168815-168827.

41. Shi Z, Groechel TR, Jain S, Chima K, Rudovic O & Matarić MJ (2022). Toward personalized affect-aware socially assistive robot tutors for long-term interventions with children with autism. *ACM Transactions on Human-Robot Interaction*, *11*(4), 1-28.

42. Vernon TW, Holden AN, Barrett AC, Bradshaw J, Ko JA, McGarry ES, Horowitz EJ, Tagavi DM, German TC (2019). A pilot randomized clinical trial of an enhanced pivotal response treatment approach for young children with autism: The PRISM model. *J Autism Dev Disord*.,49:2358–2373. doi: 10.1007/s10803-019-03909-1.

43. Hardan AY, Gengoux GW, Berquis, KL, Libove RA, Ardel CM, Phillips JM, Frazier TW & Minjarez MB (2015). A randomized controlled trial of Pivotal Response Treatment Group for parents of children with autism. *Journal of child psychology and psychiatry, and allied disciplines*, 56 8, 884-92.

44. Mohammadzaheri F, Koegel LK, Rezaee M, Rafiee SM (2014). A randomized clinical trial comparison between pivotal response treatment (PRT) and structured applied behavior analysis (ABA) intervention for children with autism. *J Autism Dev Disord*., 44:2769–2777. doi: 10.1007/s10803-014-2137-3.

45. Tapus A, Peca A, Aly A, Pop C, Jisa L, Pintea S, Rusu AS & David DO (2012). Children with autism social engagement in interaction with Nao, an imitative robot: A series of single case experiments. *Interaction Studies*. 2012. Jan 1;13(3):315–47.

46. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, Stadnick N, Zheng K, Mukamel D, Sorkin DH (2020). Barriers and facilitators to user engagement with digital mental health interventions: a systematic review. *J. Med. Internet Res.,* 23, e24387. doi: 10.2196/preprints.24387.

47. Costescu C, Podina IR & Voinescu A (2022). Digital mental health: Interventions and assessment. *Frontiers in Psychology*, 13, 1014527.

48. Esteban PG, Baxter P, Belpaeme T, Billing E, Cai H, Cao HL, Coeckelbergh M, Costescu C, David D, De Beir A, Kennedy J, Liu H, Mazel A, Pandey A, Richardson K, Senft E, Thill S, Van de Perre G, Vanderborght B, Vernon D, Yu H & Ziemke T (2017). How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn – Journal of Behavioral Robotics*, 8(1), 18-38.

49. Cai H, Fang Y, Ju Z, Costescu C, David DO, Billing E, Ziemke T, Thill S, Belpaeme T, Vanderborght B, Vernon D, Richardson K & Liu H (2018). Sensing-enhanced therapy system for assessing children with autism spectrum disorders: a feasibility study. *IEEE Sensors*, 19, 4, 1508-1518.

50. Flight L & Julious SA (2016). Practical guide to sample size calculations: an introduction. *Pharmaceutical statistics, 15*(1), 68-74.

51. Zuur AF, Ieno EN, Walker NJ, Saveliev AA & Smith GM (2009). *Mixed effects models and extensions in ecology with R*. New York: Springer.

52. Feingold A (2015). Confidence interval estimation for standardized effect sizes in multilevel and latent growth modeling. *Journal of consulting and clinical psychology, 83*(1), 157.

53. Mascha EJ & Sessler DI (2011). Equivalence and noninferiority testing in regression models and repeated-measures designs. *Anesthesia & Analgesia, 112*(3): 678-687.

54. Enders CK (2022). *Applied missing data analysis*. Guilford Publications.

55. Siddique J, Brown CH, Hedeker D, Duan N, Gibbons RD, Miranda J & Lavori PW (2008). Missing data in longitudinal trials–part B, analytic issues. *Psychiatric annals, 38*(12), 793–801.

Ghiglino, D., Floris, F., De Tommaso, D., Kompatsiari, K., Chevalier, P., Priolo, T., & Wykowska, A. (2023). Artificial scaffolding: Augmenting social cognition by means of robot technology. *Autism Research*, 16(5), 997-1008. https://doi.org/10.1002/aur.2906

Acknowledgments

Funding: This work was funded by the European Commission, FP7 grant agreement 611391 (DREAM, "Development of Robot-Enhanced therapy for children with AutisM spectrum disorders", 2014-2019). T.Z. was also supported by ELLIIT, the Excellence Center at Linköping-Lund in Information Technology.

Competing interests: Authors declare that they have no competing interests.

Data, code and materials availability: the data and code are available via the links below. The complete clinical trials platform, comprising the physical design, the data analysis algorithms, intervention scripts and robot control functions, is available as open source at https://github.com/dream2020/DREAM, licensed under GNU GPL v.3. Additionally, the dataset collected for the present study, comprising close to 4000 sessions, is released publicly at https://github.com/dream2020/data. This dataset comprises secondary data extracted by the sensorised table during therapy, whereas footage of children is excluded to preserve anonymity.

Figures:

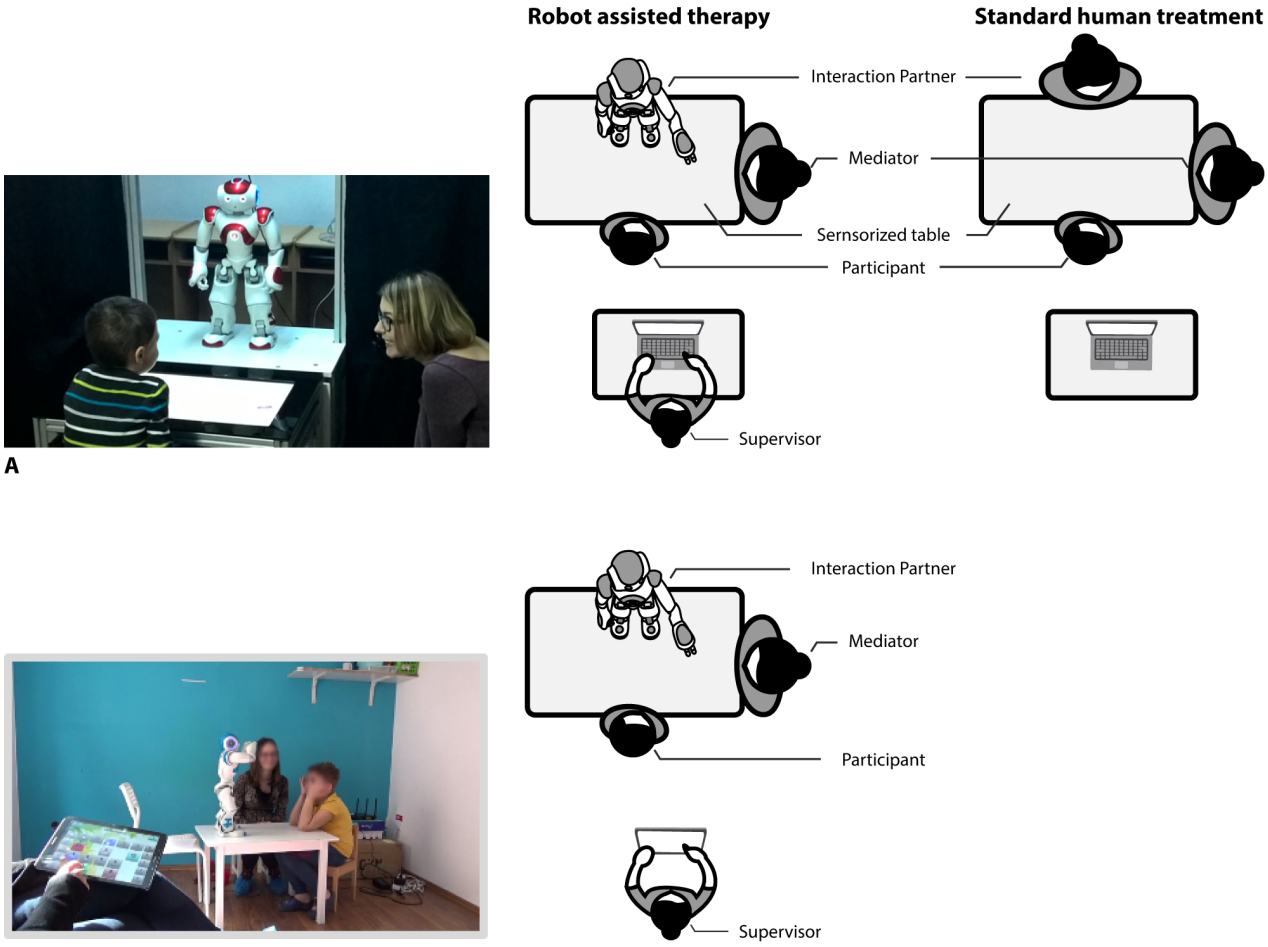


Fig. 1. Robot-assisted therapy research platforms. (**A**) Clinical trials platform used in Study 1. (**B**) educational platform used in Study 2.

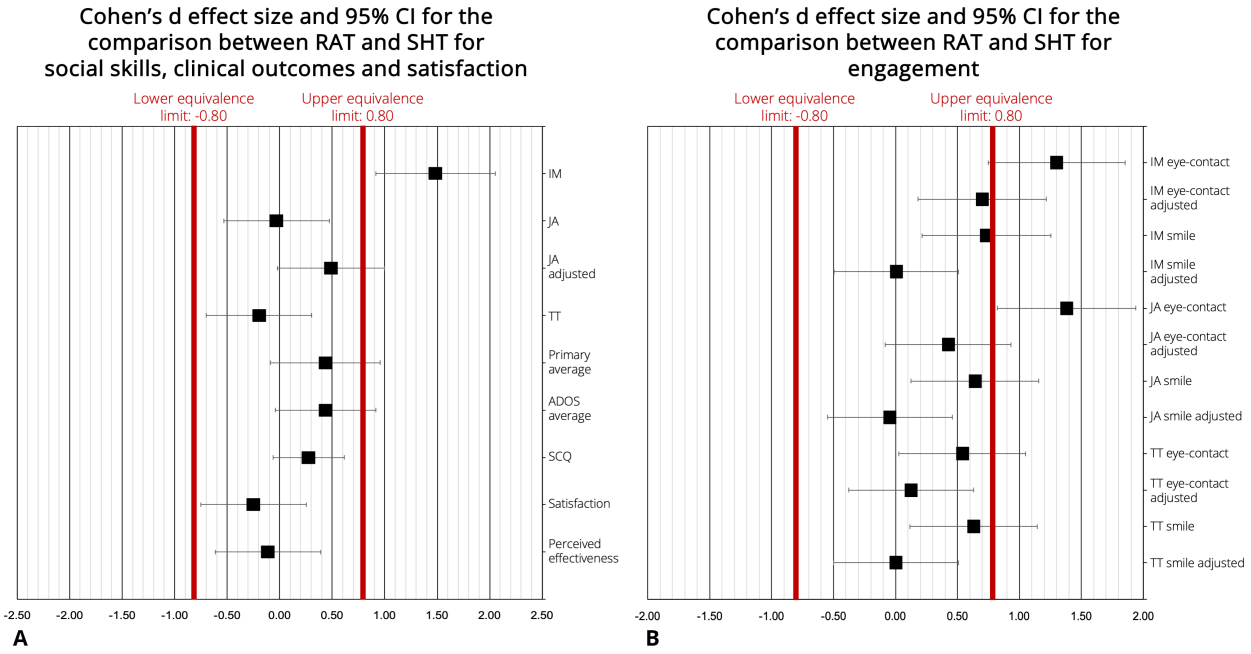


Fig. 2. Study 1 results: efficacy of robot-assisted therapy using the clinical trials platform – primary outcomes (social skills, clinical outcomes, satisfaction, and perceived effectiveness) and secondary outcomes (eye-contact and smiling, i.e., engagement). The comparison of RAT to SHT on social skills, clinical outcomes, and satisfaction (A) using Cohen's *d* effect sizes and their 95% confidence intervals indicate a significant difference between RAT and SHT for imitation, d = 1.48, 95% confidence interval [0.91; 2.05]. All other assessed criteria fell within equivalence limits. The comparison of RAT to SHT on engagement (B) using Cohen's *d* effect sizes and their 95% confidence intervals indicate a significant difference between RAT and SHT on eye-contact in the joint-attention task when the scores are not adjusted for baseline differences, *d* = 1.38, 95% confidence interval [0.83; 1.93]. All other assessed criteria fell within equivalence limits. Cohen's *d* effect sizes for all comparisons were computed using the estimated marginal means and their standard errors from a hierarchical linear model.

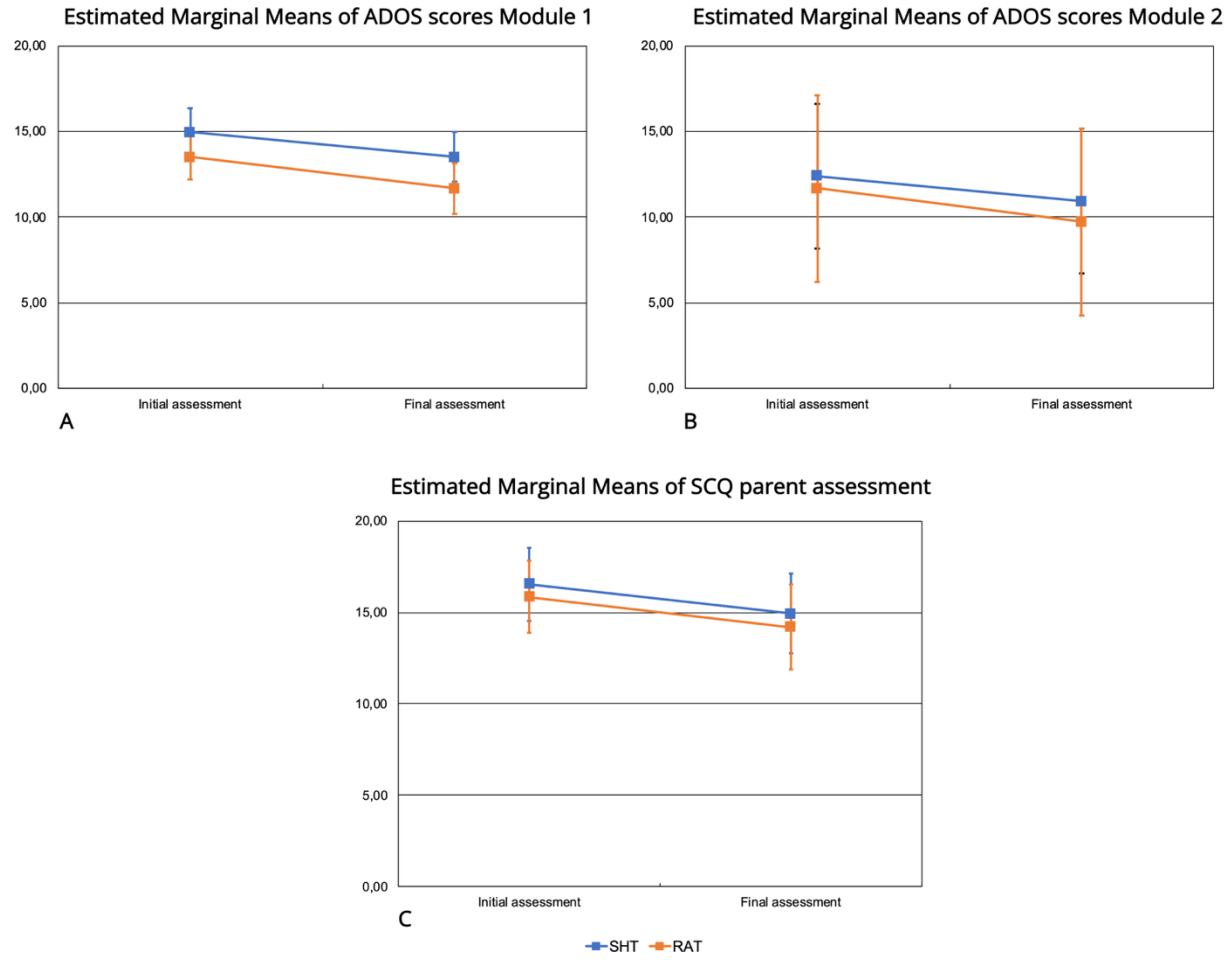


Fig. 3. Study 1 results: efficacy of robot-assisted therapy using the clinical trials platform – clinical outcomes (autism disorder observation schedule ADOS and social communication questionnaire SCQ). ADOS scores for Module 1 (A) were analysed using an hierarchical linear model that included fixed effects for time (pre- vs. post-test), group, group * time interaction, and a fixed intercept. Results indicate a significant effect for time ($F(1, 39.75) = 21.46$, $p < 0.001$). No significant effects for group ($F(1, 60.75) = 3.15$, $p = 0.381$) or interaction ($F(1, 39.75) = 0.30$, $p = 0.586$) were found. The estimated marginal means for the time effect indicated that scores have decreased from pre-test (mean = 14.23, standard error = 0.48) to post-test (mean = 12.60, standard error = 0.52). SCQ parent assessment (C) were analysed using a similar hierarchical linear model including fixed effects: intercept, time (pre- vs. post-test), group, and group * time interaction. The results indicated a significant effect for time ($F(1, 37.22) = 7.80$, $p = 0.008$), but no significant effects of group ($F(1, 69.81) = 0.27$, $p = 0.607$) or interaction ($F(1, 37.22) = 0.01$, $p = 0.966$). The estimated marginal means for the time effect indicated that scores have decreased from pre-test (mean = 16.20, standard error = 0.70) to post-test (mean = 14.56, standard error = 0.81). ADOS scores for Module 2 (B) were not analysed statistically due to low statistical power, see supplementary material for details.

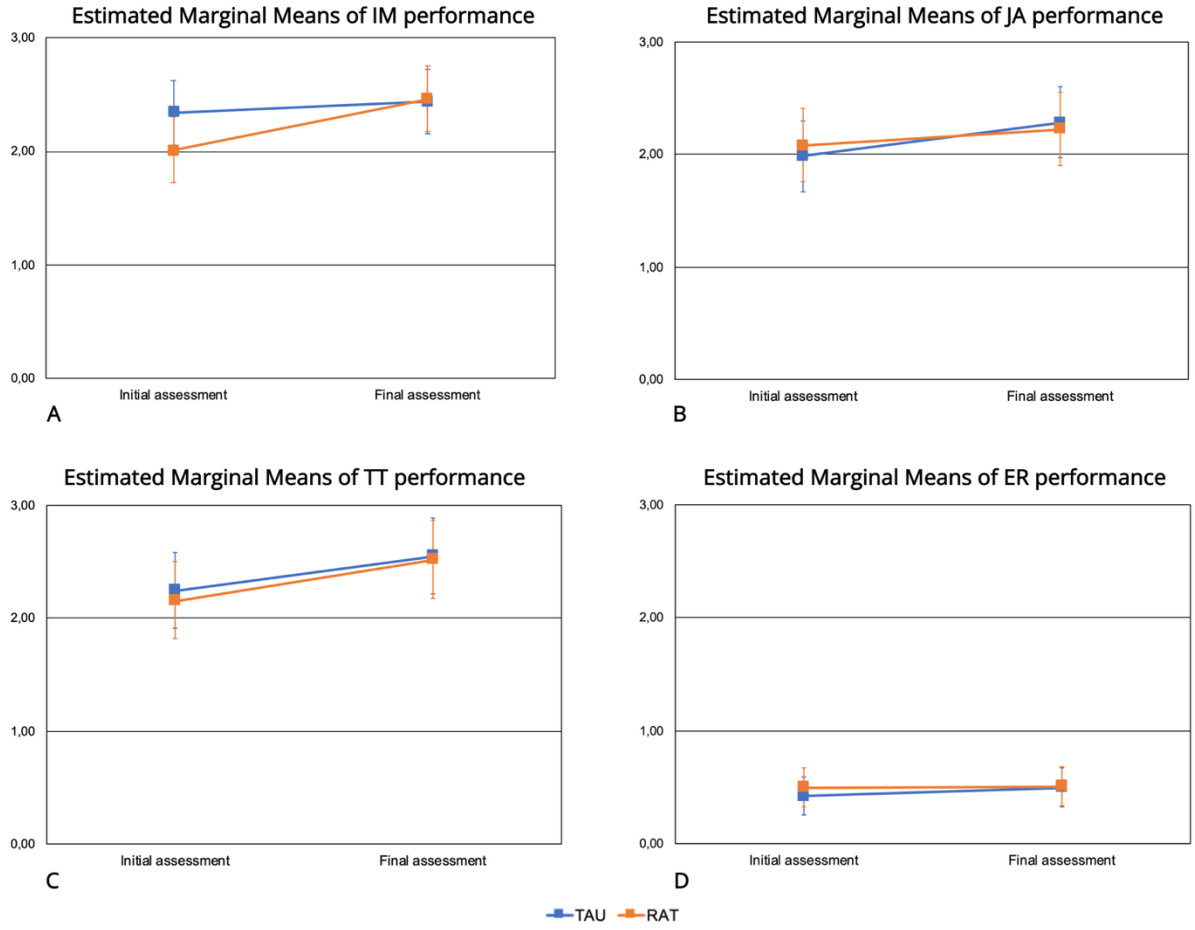


Fig. 4. Study 2 results: effectiveness of robot-assisted therapy using the educational platform – performance in imitation, joint attention, turn-taking, and emotion recognition.

All four measures (imitation, joint attention, turn-taking, and emotion recognition) were analysed using hierarchical linear models including a fixed intercept, and fixed effects for time (pre- vs. post-test), group, and group * time interaction. For imitation (A), significant effect for time ($F(1, 56.34) = 21.25$, $p < 0.001$) and interaction ($F(1, 56.34) = 8.97$, $p = 0.004$) were found, but the group effect was not significant ($F(1, 61.44) = 0.64$, $p = 0.427$). The estimated marginal means for the time effect indicated that scores have increased from pre-test (mean= 2.18, standard error = 0.10) to post-test (mean = 2.45, standard error = 0.10). The estimated marginal means for the interaction effect indicated that the scores have increased significantly in the RAT group from pre-test (mean = 2.01, standard error = 0.14) to post-test (mean = 2.46, standard error = 0.15), $p < 0.001$, but no significant change were observed in the control group, $p = 0.247$. For turn-taking (C) results indicate a significant effect for time ($F(1, 58.11) = 11.63$, $p = 0.001$). The effects for group ($F(1, 60.77) = 0.07$, $p = 0.789$) and interaction ($F(1, 58.11) = 0.08$, $p = 0.775$) were not significant. The estimated marginal means for the time effect indicated that scores have increased from pre-test (mean = 2.20, standard error = 0.12) to post-test (mean = 2.53, standard error = 0.12). No significant differences for joint attention (B) and emotion recognition (D) were found.

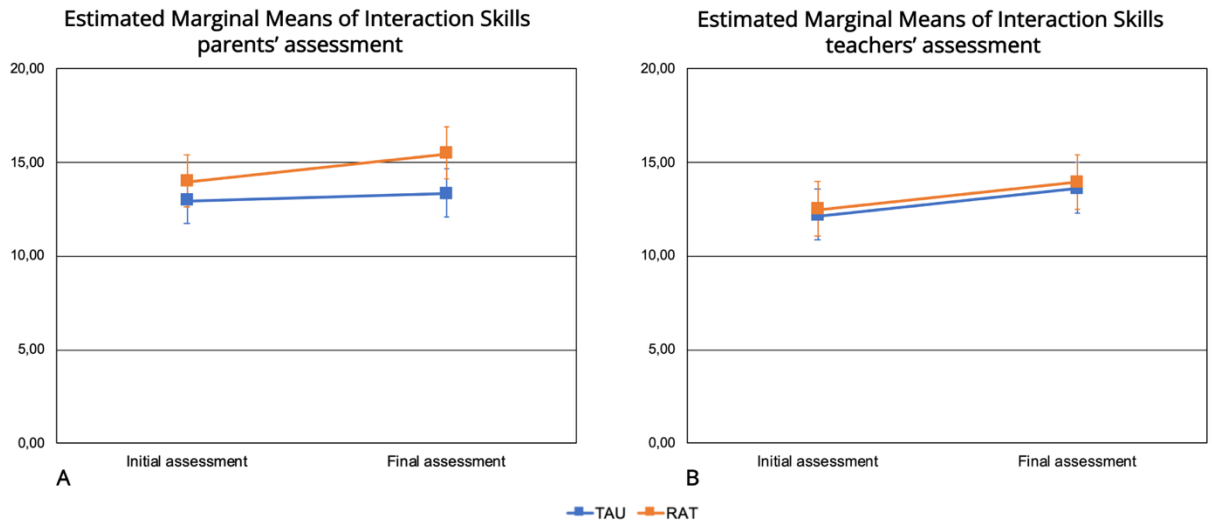


Fig. 5.Study 2 results: effectiveness of robot-assisted therapy using the educational platform – Performance in social interaction assessed by parents and teachers.. Interaction skill (IS) assessments (by both parents and teachers) were analysed using hierarchical linear models including fixed effects for the intercept, time (pre- vs. post-test), group, and group * time interaction. Results for parents' assessments (A) indicated a significant effect for time ($F(1, 51.52) = 4.88$, $p = 0.032$) but not for group ($F(1, 57.17) = 3.60$, $p = 0.063$) or interaction ($F(1, 51.52) = 1.742$, $p = 0.193$). The estimated marginal means for the time effect indicate that scores have increased from pre-test (mean = 13.45, standard error = 0.45) to post-test (mean =14.38, standard error = 0.48). Results for teachers' assessments (B) indicate a significant effect for time ($F(1, 53.86) = 12.14$, $p < 0.001$) but not for group ($F(1, 57.67) = 0.001$, $p = 0.971$) or interaction ($F(1, 53.86) = 0.52$, $p = 0.472$). The EMMs for the time effect indicated that scores have increased from pre-test (mean = 12.27, standard error = 0.49) to post-test (mean =13.46, standard error = 0.49).

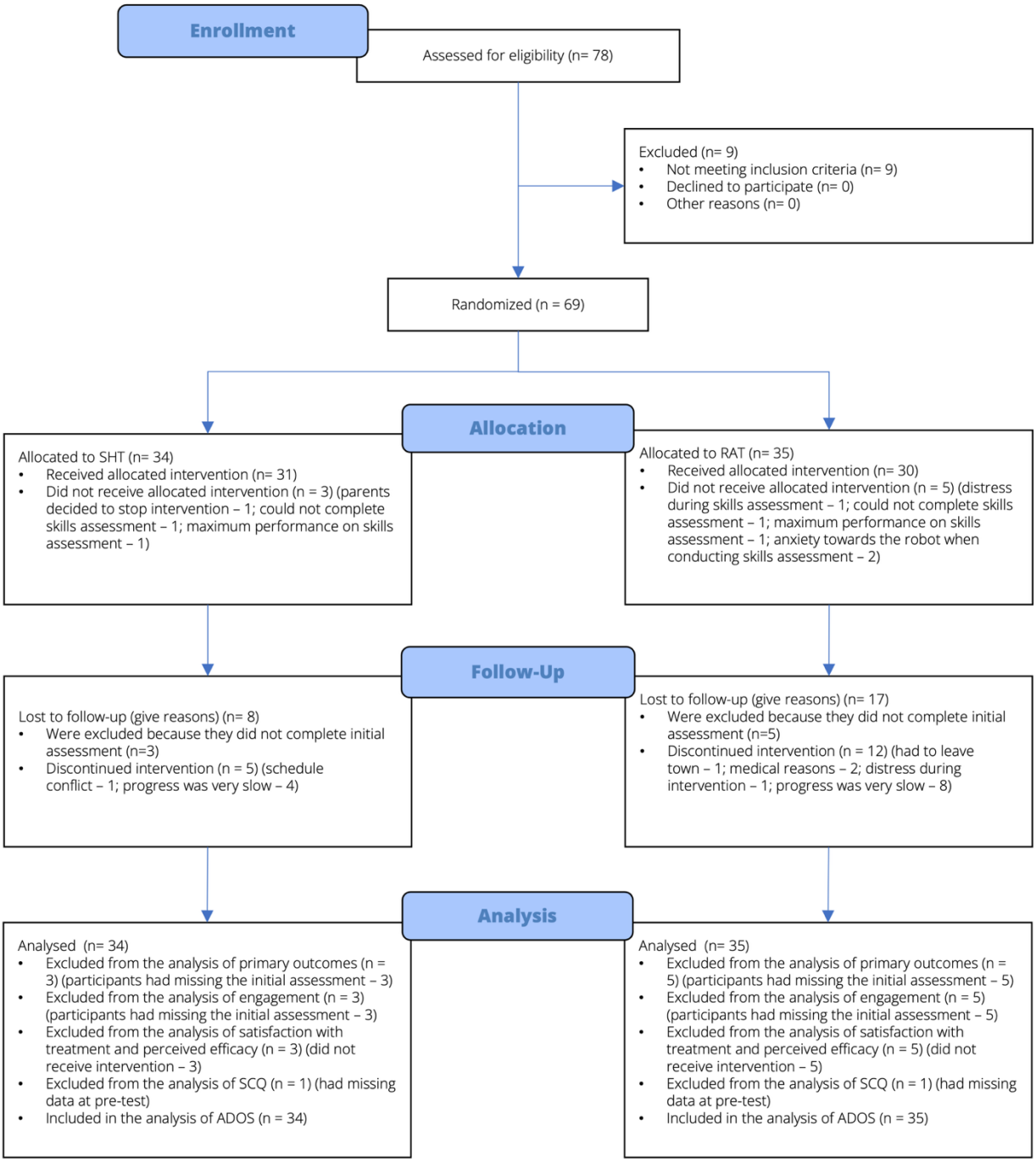


Fig. 6. Study 1: Efficacy of robot-assisted therapy using the clinical trials platform. CONSORT Flow Diagram showing how participants were enrolled, allocated to the SHT and RAT groups, excluded or discontinued, and analysed.

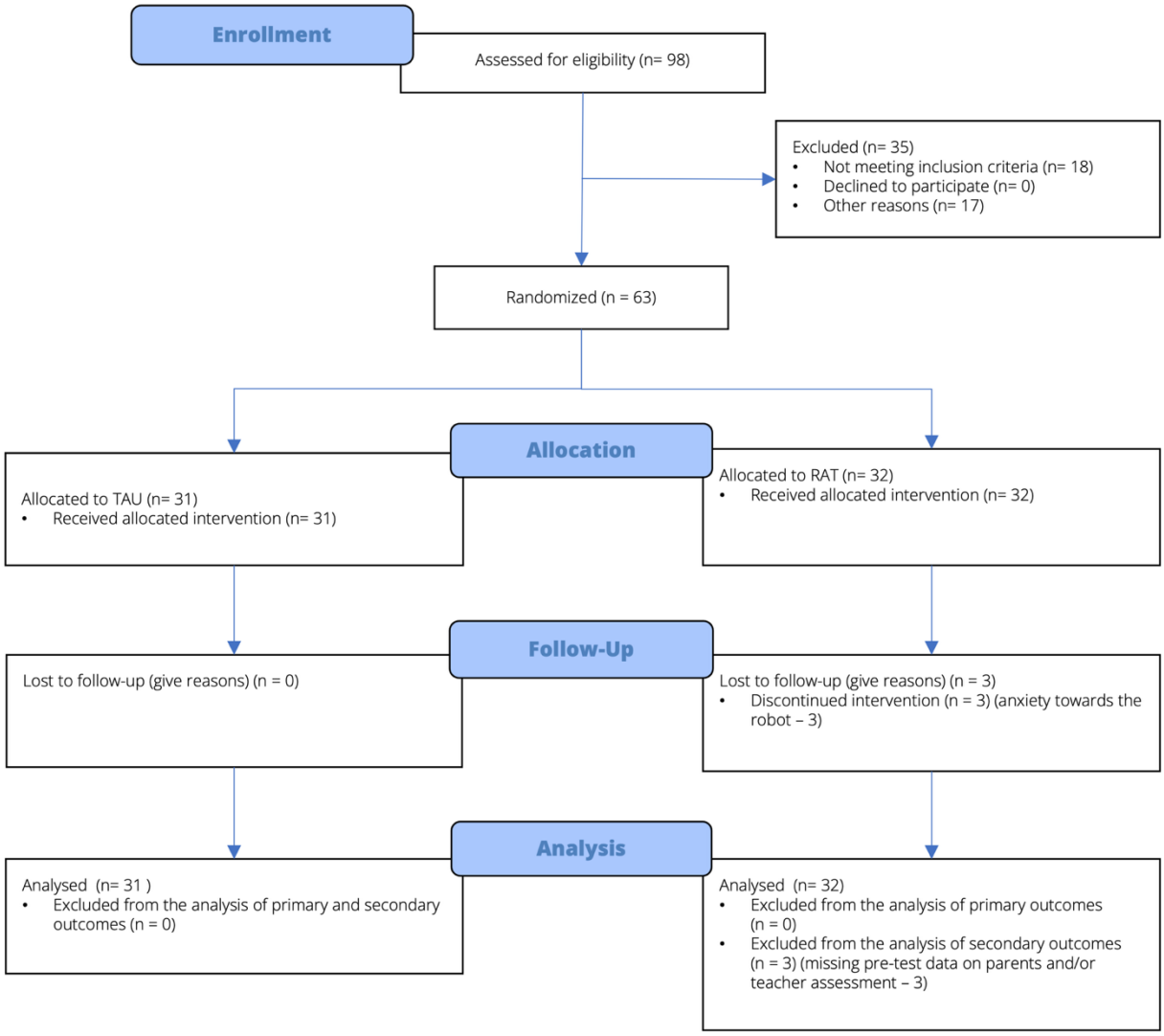


Fig. 7. Study 2: Effectiveness of robot-assisted therapy using the educational platform. CONSORT Flow Diagram. showing how participants were enrolled, allocated to the treatment as usual group and RAT group, excluded or discontinued, and analysed.